Grid Computing Inside Hospitals Using Virtualization Technology: A Secure Solution For Heavy Computing Tasks

Xin Zhou, Adrien Depeursinge, Marko Niinimaki, Antoine Geissbuhler and Henning Müller
Geneva University Hospitals and University of Geneva, Geneva, Switzerland

Contact address: Xin Zhou, Service for medical informatics, University Hospitals of Geneva, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva 14, Switzerland. Email: xin.zhou@sim.hcuge.ch

Introduction

Challenges

• The **ever-increasing** production of medical images in digital format calls upon new tools for the data analysis to use the data up to their full potential. Much of the data analysis and automatic processing of large image datasets can be **computationally expensive**.

• Like most hospitals, the University Hospitals of Geneva (HUG) do not have any central research computing infrastructure to execute computationally intensive applications at the moment. Using external computing resources can cause **legal problems** due to the data transfer of patient data.

Solution

• Over **6'000** desktop PCs are available on the network of the HUG. Even a **partial re-use** of these resources could help to fulfill the researchers' computational needs.

• To **explore** the **idle** computing resources, **virtualization techniques (VMware)** are deployed. The whole infrastructure is built up based on the **ARC (Advanced Resource Connector)** middleware.

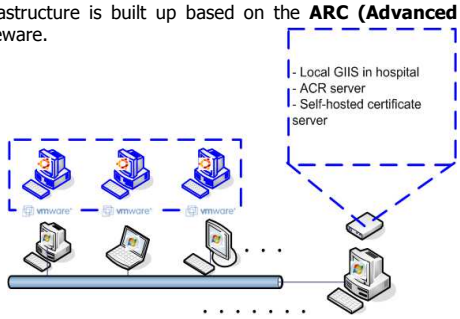

Figure 1. Infrastructure of internal Grid

Results

System deployment

• **Fully automatic installation** through the HUG network (standard solution of the HUG based on Microsoft active directory). The infrastructure is described in Figure 1.

• Computing node requirements : free hard disk > 2GB, RAM > 768MB. Server requires more disk space, can be ranged with an external hard disk.

Performance comparison

• Test application consisted of 50 jobs, each job treated 1,000 images.

• Total time results is shown in Table 1, statistics of execution time for each job is shown in Figure 2.

Table 1. Total execution time comparison

	Time (min)
1 server with 4 CPU, 4GB RAM	807
Remote Grid with 37 CPU	537
Local Grid with 10CPU, 350M x10 RAM	240

Disturbance measurement

• Disturbance of the working environment and the HUG network is negligible (See Table 2). Supposing that the network capacity is 100 Mbit/s and congestion is at 75% (~ 8 MB/s), 1 VM client would use 0.05% of the network capacity. The administrator can limit the network debit of the Grid, and users can stop the virtual system to take resources back.

• The impact is tested with frequently used applications (See Table 3).

Table 2. Disturbance of the working environment and the HUG network

	PC idle, VM not running	PC idle, VM started	ProVision and browser in use	Grid running with 25M input file
Measurement during 10 minutes				
bytes w/o job files	29,000	59,800	4,500,000	61,000
job files				26,009,600
bytes/s	48	100	7,500	43,451
bits/s	387	797	60,000	347,608
Measurement during 1 hour				
bytes/s	76	105	7,300~15,000	27,290
bits/s	608	840	61,600	218,320

Table 3. Office applications start up time

	VM not running	VM running but idle	VM running Grid jobs
Old machine (CPU 2.8GHz, 768MB RAM)			
Internet Explorer 7	2.7	3.2	5.8
MS office 2003	2.8	3.3	8.1
New machines (Dual-core 2.8GHz, 2GB RAM), All machines will be upgraded to this configuration by mid 2009.			
Internet Explorer 7	2.7	2.7	2.8
MS office 2003	2.4	2.5	3.0
MS office 2007	4.2	4.5	6.1

Methods & Materials

Hardware

• **20** standard (old) **desktop** PCs (with 2.8GHz CPU and 768M RAM) of the HUG are used as a small **test bed** for our intra-hospital Grid.

Software

• A virtual Linux operating system was installed using VMWare on each PC. This **virtual machine (VM)** serves as a computing node, which separates the computation node from the user's operating system.

Evaluation

• The system is evaluated based on the **usability** for research purposes and the possible **disturbance** of the HUG network, for example through data transfers and through slowing the desktop system.

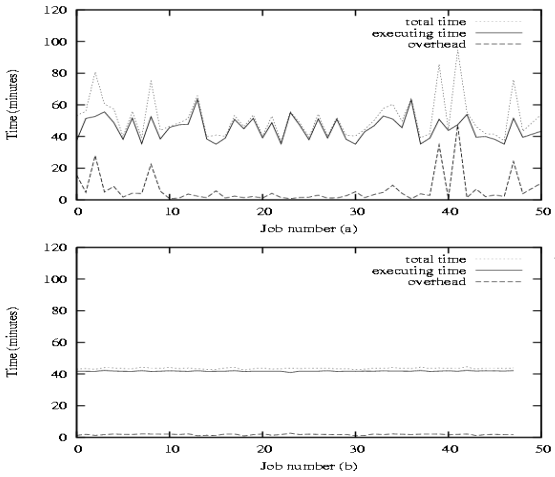

Figure 2. Detail Grid execution time evaluation.

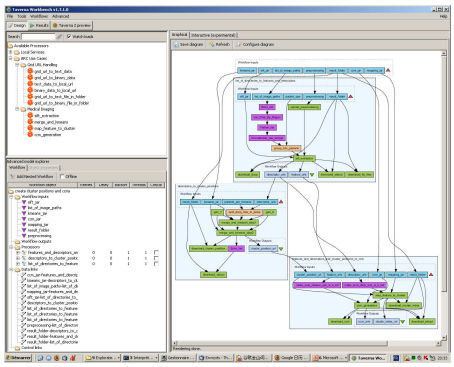

Figure 3. A screenshot of the graphical tool to develop (or adapt) Grid-based applications.

Conclusions

Now available

• **An internal computation Grid** containing 20 PCs supports researchers with computing power.

• **An internal certificate server** allows to distribute personal certificates valid inside hospital. Short Lived Credential Service (SLCS) can provide temporal certificates valid for the whole Switzerland.

• Tools with **user-friendly** Graphical User Interface (GUI) help users to quickly adapt Grid computing without much knowing technical details (See Figure 3).

• Three medical imaging applications (general content-based image retrieval, lung image retrieval, and fracture image retrieval) were griddified.

Future work

• **Extend** the internal Grid based on volunteers (mainly researchers).

• Open for new applications.

Acknowledgements

• This work was funded by the European Union 6th Framework program in the context of the **KnowARC research project**. We would also like to thank the **DIG** for their support throughout the project and the quick responses we often received.